Evolutionary Dynamics of the Leucine-Rich Repeat Receptor-Like Kinase (LRR-RLK) Subfamily in Angiosperms[1][OPEN]

Iris Fischer*, Anne Diévart, Gaetan Droc, Jean-François Dufayard, and Nathalie Chantret*

Institut National de la Recherche Agronomique, Unité Mixte de Recherche Amélioration Génétique et Adaptation des Plantes Méditerranéennes et Tropicales, F–34060 Montpellier, France (I.F., N.C.); and Centre de Coopération Internationale en Recherche Agronomique Pour le Développement, Unité Mixte de Recherche AGAP, F–34398 Montpellier, France (A.D., G.D., J.-F.D.)

ORCID IDs: 0000-0002-5080-1223 (I.F.); 0000-0001-9460-4638 (A.D.); 0000-0003-1849-1269 (G.D.).

Gene duplications are an important factor in plant evolution, and lineage-specific expanded (LSE) genes are of particular interest. Receptor-like kinases expanded massively in land plants, and leucine-rich repeat receptor-like kinases (LRR-RLK) constitute the largest receptor-like kinases family. Based on the phylogeny of 7,554 LRR-RLK genes from 31 fully sequenced flowering plant genomes, the complex evolutionary dynamics of this family was characterized in depth. We studied the involvement of selection during the expansion of this family among angiosperms. LRR-RLK subgroups harbor extremely contrasting rates of duplication, retention, or loss, and LSE copies are predominantly found in subgroups involved in environmental interactions. Expansion rates also differ significantly depending on the time when rounds of expansion or loss occurred on the angiosperm phylogenetic tree. Finally, using a d_N/d_S-based test in a phylogenetic framework, we searched for selection footprints on LSE and single-copy LRR-RLK genes. Selective constraint appeared to be globally relaxed at LSE genes, and codons under positive selection were detected in 50% of them. Moreover, the leucine-rich repeat domains, and specifically four amino acids in them, were found to be the main targets of positive selection. Here, we provide an extensive overview of the expansion and evolution of this very large gene family.

Receptor-like kinases (RLKs) constitute one of the largest gene families in plants and expanded massively in land plants (Embryophyta; Lehti-Shiu et al., 2009, 2012). For plant RLK gene families, the functions of most members are often not known (especially in recently expanded families), but some described functions include innate immunity (Albert et al., 2010), pathogen response (Dodds and Rathjen, 2010), abiotic stress (Yang et al., 2010), development (De Smet et al., 2009), and sometimes multiple functions (Lehti-Shiu et al., 2012). The RLKs usually consist of three domains: an N-terminal extracellular domain, a transmembrane domain, and a C-terminal kinase domain (KD). In plants, the KD usually has a Ser/Thr specificity (Shiu and Bleecker, 2001), but Tyr-specific RLKs were also described (e.g. BRASSINOSTEROID INSENSITIVE1; Oh et al., 2009). Interestingly, it was estimated that approximately 20% of RLKs contain a catalytically inactive KD (e.g. STRUBBELIG and CORYNE; Chevalier et al., 2005; Castells and Casacuberta, 2007; Gish and Clark, 2011). In Arabidopsis (*Arabidopsis thaliana*), 44 RLK subgroups (SGs) were defined by inferring the phylogenetic relationships between the KDs (Shiu and Bleecker, 2001). Interestingly, different SGs show different duplication/retention rates (Lehti-Shiu et al., 2009). Specifically, RLKs involved in stress responses show a high number of tandemly duplicated genes whereas those involved in development do not (Shiu et al., 2004), which suggests that some RLK genes are important for the responses of land plants to a changing environment (Lehti-Shiu et al., 2012). There seem to be relatively few RLK pseudogenes compared with other large gene families, and copy retention was argued to be driven by both drift and selection (Zou et al., 2009; Lehti-Shiu et al., 2012). As most SGs are relatively old and RLK subfamilies expanded independently in several plant lineages, duplicate retention cannot be explained by drift alone, and natural selection is expected to be an important driving factor in RLK gene family retention (Lehti-Shiu et al., 2009).

[1] This work was supported by the German Research Foundation (grant no. FI 1984/1–1 to I.F.), by the Agropolis Resource Center for Crop Conservation, Adaptation, and Diversity project of the Agropolis Foundation (to I.F.), and by the Agence Nationale de la Recherche (grant no. ANR–08–GENM–021).

* Address correspondence to irisfischer402@gmail.com and chantret@supagro.inra.fr.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Nathalie Chantret (chantret@supagro.inra.fr).

I.F., A.D., and N.C. designed the study; G.D. performed the LRR-RLK extraction; J.-F.D. performed the phylogenetic clustering; I.F., A.D., and N.C. performed the data and statistical analyses; I.F. wrote the article with the help of A.D. and N.C.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.01470

Leucine-rich repeat-receptor-like kinases (LRR-RLKs), which contain up to 30 leucine-rich repeat (LRRs) in their extracellular domain, constitute the largest RLK family (Shiu and Bleecker, 2001). Based on the KD, 15 LRR-RLK SGs have been established in Arabidopsis (Shiu et al., 2004; Lehti-Shiu et al., 2009). So far, two major functions have been attributed to them: defense against pathogens and development (Tang et al., 2010b). LRR-RLKs involved in defense are predominantly found in lineage-specific expanded (LSE) gene clusters, whereas LRR-RLKs involved in development are mostly found in nonexpanded groups (Tang et al., 2010b). It was also discovered that the LRR domains are significantly less conserved than the remaining domains of the LRR-RLK genes (Tang et al., 2010b). In addition, a study of four plant genomes (Arabidopsis, grape [*Vitis vinifera*], poplar [*Populus trichocarpa*], and rice [*Oryza sativa*]) showed that LRR-RLK genes from LSE gene clusters show significantly more indications of positive selection or relaxed constraint than LRR-RLKs from nonexpanded groups (Tang et al., 2010b).

The genomes of flowering plants (angiosperms) have been shown to be highly dynamic compared with most other groups of land plants (Leitch and Leitch, 2012). This dynamic is mostly caused by the frequent multiplication of genetic material, followed by a complex pattern of differential losses (i.e. the fragmentation process) and chromosomal rearrangements (Langham et al., 2004; Leitch and Leitch, 2012). Most angiosperm genomes sequenced so far show evidence for at least one whole-genome multiplication event during their evolution (Jaillon et al., 2007; D'Hont et al., 2012; Tomato Genome Consortium, 2012). At a smaller scale, tandem and segmental duplications are also very common in angiosperms (Arabidopsis Genome Initiative, 2000; International Rice Genome Sequencing Project, 2005; Rizzon et al., 2006). Although the most common fate of duplicated genes is to be progressively lost, in some cases they can be retained in the genome, and adaptive as well as nonadaptive scenarios have been discussed to play a role in this preservation process (for review, see Moore and Purugganan, 2005; Hahn, 2009; Innan, 2009; Innan and Kondrashov, 2010). Whole-genome sequences also revealed that the same gene may undergo several rounds of duplication and retention. These LSE genes were shown to evolve under positive selection more frequently than single-copy genes in angiosperms (Fischer et al., 2014). That study analyzed general trends over whole genomes. Here, we ask if, and to what extent, this trend is observable at LRR-RLK genes. As this gene family is very dynamic and large, and in accordance with the results of Tang et al. (2010b), we expect the effect of positive selection to be even more pronounced than in the whole-genome average.

We analyzed 33 Embryophyta genomes to investigate the evolutionary history of the LRR-RLK gene family in a phylogenetic framework. Twenty LRR-RLK SGs were identified, and from this data set, we deciphered the evolutionary dynamics of this family within angiosperms. The expansion/reduction rates were contrasted between SGs and species as well as in ancestral branches of the angiosperm phylogeny. We then focused on genes whose number increased dramatically in an SG- and/or species-specific manner (i.e. LSE genes). Those genes are likely to be involved in species-specific cellular processes or adaptive interactions and were used as a template to infer the potential occurrence of positive selection. This led to the identification of sites at which positive selection likely acted. We discuss our results in the light of angiosperm genome evolution and current knowledge of LRR-RLK functions. Positive selection footprints identified in LSE genes highlight the importance of combining evolutionary analysis and functional knowledge to guide further investigations.

RESULTS

We extracted genes containing both LRRs and a KD from 33 published embryophyte genomes. Here, we mostly describe the findings for the 31 angiosperm (eight monocot and 23 dicot) genomes we analyzed. The 7,554 LRR-RLK genes were classified in 20 SGs. This classification was inferred using distance-related methods, because the high number of sequences to be analyzed would imply excessive computation time for methods relying on maximum likelihood. Since we decided to study the evolutionary dynamic of the LRR-RLK gene family using SG classification as a starting point, we first wanted to verify that each SG was monophyletic. Ten subsets of about 750 sequences were created by picking one sequence out of 10 to infer a PHYML tree (data not shown). Analysis of the trees shows that most SGs (14) are monophyletic with strong branch support. On the other hand, for six SGs (SG_I, SG_III, SG_VI, SG_Xb, SG_XI, and SG_XV), the topology differs slightly between trees: in at least five trees out of 10, either the SG appears to be paraphyletic or few sequences are placed outside the main monophyletic clade with low branch support. As we could not confirm that these SGs are monophyletic, they were tagged with an asterisk throughout this article.

Next, we determined the number of ancestral genes present in the last common ancestor of angiosperms (LCAA) using a tree reconciliation approach (see "Materials and Methods"). In short, tree reconciliation compares each SG-specific LRR-RLK gene tree with the species tree to infer gene duplications and losses. Note that since only LRR-RLKs with at least one complete LRR were considered, some of the inferred gene losses might correspond to RLKs without, or with degenerated, LRRs. Using this method, we predicted the number of LRR-RLK genes in the LCAA to be 150. All SGs were present in the LCAA, but the number of genes between SGs was highly variable (Table I). SG_III* and SG_XI* show the highest number of ancestral genes, with 32 and 29 genes, respectively. The lowest numbers of ancestral genes are recorded for SG_VIIb, SG_Xa, SG_XIIIa, and SG_XIIIb, which only possessed two genes, and SG_XIV, which only contained one. These results show that, already in

Table I. *Total number of LRR-RLKs in our angiosperm data set, number of ancestral genes in the LCAA, and median global expansion rate for each SG among the 31 species*

SG	Total No. of Genes	No. of Ancestral Genes	Median Global Expansion Rate
I*	482	7	2.00
II	349	9	1.22
III*	1,400	32	1.22
IV	131	3	1.33
V	263	5	1.80
VI*	324	10	1.00
VIIa	157	3	1.67
VIIb	84	2	1.50
VIII-1	216	5	1.40
VIII-2	355	8	1.25
IX	193	3	1.67
Xa	143	2	2.00
Xb*	367	9	1.11
XI*	1,177	29	1.28
XIIa	1,126	9	3.00
XIIb	423	4	2.00
XIIIa	84	2	1.50
XIIIb	77	2	1.00
XIV	84	1	3.00
XV*	119	5	0.80
Total	7,554	150	

the LCAA, which lived approximately 150 million years ago (Supplemental Table S1), some SGs were more prone to retain copies than others. We wanted to determine if this ancestral pattern was preserved during the course of angiosperm evolution and if different SGs expanded or contracted compared with the LCAA.

Expansion Rates of LRR-RLK Genes Differ between Subgroups and Species

To gain a more comprehensive understanding of LRR-RLK evolution, we first looked at SG-specific expansion rates in two complementary ways. First, we calculated the global SG expansion rate (the ratio of contemporary LRR-RLK genes per species in one SG divided by the ancestral number) for each SG (Fig. 1). Second, we inferred the branch-specific expansion rate of each SG on the phylogenetic tree of the 31 angiosperm species. We did this by automatically computing the ratio of descendant LRR-RLKs divided by the ancestral number of LRR-RLKs at every node (see "Materials and Methods"; Fig. 2). Looking at the global SG expansion, we found that SG_Xa, SG_XIIa, SG_XIIb, and SG_XIV expanded more than 2-fold on average, and SG_I* and SG_IX expanded around 2-fold (Fig. 1; Supplemental Table S2). Interestingly, SG_XIIa already had a moderately high ancestral gene number (nine) and, therefore, seems to be generally prone to high retention rates. Indeed, SG_XIIa was subject to repeated rounds of major expansion events (i.e. expansion greater than 2-fold) during its evolutionary history (e.g. in Poaceae, the *Solanum* ancestor, Malvaceae, and the Arabidopsis ancestor) but also species-specific expansions, e.g. in THECC, GOSRA, ARALY, SCHPA, MALDO, LOTJA, POPTR, and JATCU (Fig. 2; for five-digit species codes, see Table II). On the other hand, SG_I* and SG_XIIb had a medium number of copies in the ancestral genome (seven and four, respectively) but the pattern of expansion is quite different when analyzed in detail (Fig. 2). For SG_I*, the expansion rate is mostly due to ancestral expansion events rather than species-specific ones. For example, the high number of copies in ARATH and EUTSA (Fig. 1) is not due to expansions specific to these species but rather an expansion in Brassicaceae. Subsequently, copies were lost in the other species of this family analyzed here (ARALY, SCHPA, BRARA) but retained in ARATH and EUTSA (Fig. 2). Species-specific expansions can also be observed in SG_I*, mostly in PRUPE and POPTR. For SG_XIIb, on the other hand, the high expansion rate is mostly due to recent species-specific expansions in PHODC, MUSAC, VITVI, GOSRA, MALDO, POPTR, and JATCU. But one major ancestral expansion can be observed in Rosids.

SG_IX, SG_Xa, and SG_XIV had only a few copies in the LCAA (three, two, and one, respectively), and all show a relatively high global expansion rate (Fig. 1). For these SGs also, a contrasted branch-specific expansion pattern can be observed (Fig. 2). SG_Xa went through relatively few major expansions: one can be detected in the dicot ancestor and a species-specific one in POPTR. Likewise, SG_IX shows only one ancestral expansion in Malvaceae but more species-specific expansions in PHODC, MUSAC, MALDO, and GLYMA. Finally, SG_XIV went through several rounds of ancestral (monocots, dicots, Malvaceae, and Brassicaceae) as well as species-specific (PHODC, MALDO, and POPTR) expansion. The other SGs show a moderate expansion rate (1.3–1.75) or no expansion at all (Fig. 1; Supplemental Table S2). SG_XV* is the only SG for which the number of copies was decreasing on average compared with the LCAA genome (0.77). It is important to note that the LCAA ancestral gene number could have been overestimated slightly for those SGs without a confirmed monophyletic origin (denoted by asterisks), resulting in an underestimation of global expansion rate. However, we recalculated the global expansion rates for each of those SGs using the largest subset of sequences that always include a stable monophyletic clade. The obtained global expansion rate differed only slightly from the ones presented here (data not shown), and the conclusions drawn remain unchanged.

Because some species underwent whole-genome duplication (WGD) or whole-genome triplication (WGT) relatively recently compared with others (Table III), we determined species-specific patterns of LRR-RLK expansions and determined if those patterns are consistent with the recent history of the species. Therefore, we computed the global species expansion rate (the ratio of LRR-RLK genes per SG in one species divided by the ancestral number) for each of the 31 angiosperm species. As expected, the global expansion rate differs significantly between species (Fig. 3; Supplemental Table S2). Compared with the LCAA (150 genes), the number of LRR-RLK genes did not

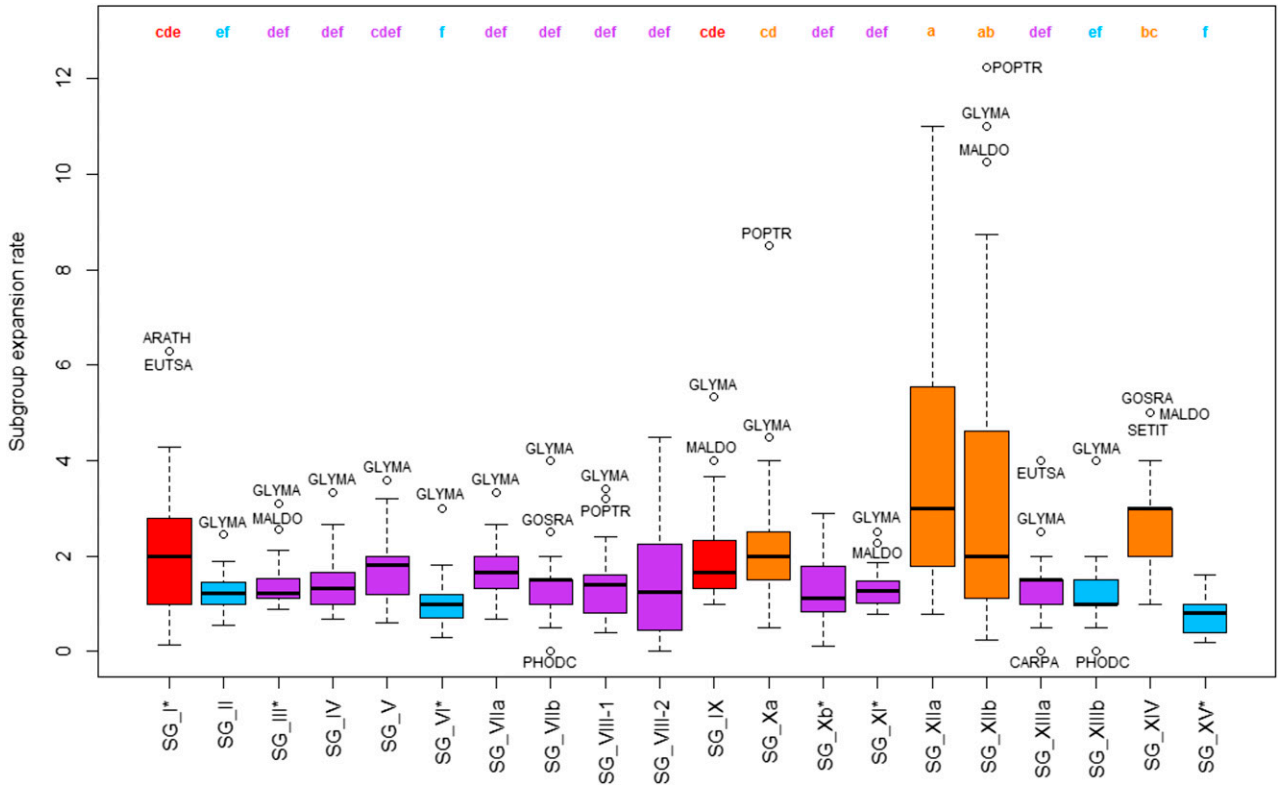

Figure 1. Global expansion rate in each SG, which is the total number of genes in each species divided by the ancestral number (Table I). An ANOVA test showed that the expansion rate differs significantly between SGs ($P < 2e\text{-}16$). Therefore, we performed a TukeyHSD test to determine which SGs exactly show a significant difference between each other and grouped those SGs by significance level (a–e). Letters above each box plot indicate the TukeyHSD significance group (Supplemental Table S2). The significance groups are color coded according to the mean expansion rate: orange, greater than 2.25-fold expansion; red, 1.75- to 2.25-fold expansion; purple, 1.3- to 1.75-fold expansion; and blue, 0.75- to 1.3-fold expansion (i.e. no expansion). The outlier species are labeled for each SG. For species identifiers, see Table II.

decrease for most species except for LOTJA (114) and CARPA (127). This indicates that, on average, LRR-RLK genes are more prone to retention than loss. Some species, however, did not significantly expand their average number of LRR-RLK genes compared with the common ancestor: PHODC (158), CUCME (149), CUCSA (180), SCHPA (194), BRARA (185), MEDTR (183), and RICCO (182). LRR-RLK genes expanded more than 2-fold in GLYMA (477), MALDO (441), POPTR (400), and GOSRA (372) and around 2-fold in MUSAC (280), MAIZE (241), SETIT (301), ORYSJ (317), ORYSI (301), SOLTU (254), PRUPE (260), MANES (238), and EUTSA (240). The remaining species show a moderate expansion rate (1.4–1.75): CACJA (222), THECC (238), JATCU (208), ARATH (222), SOLLC (232), SORBI (225), BRADI (225), VITVI (193), and ARALY (195). As expected, the four species with the highest global expansion rate (GLYMA, MALDO, POPTR, and GOSRA) are recent polyploids in which most SGs have expanded (Fig. 2). However, some SGs expanded more than 2-fold, indicating that small-scale duplication events have occurred in addition to polyploidy. In POPTR, for instance, the global expansion rates of SG_Xa and SG_XIIb are more than 8-fold (Fig. 3), and a strong branch-specific expansion rate is detected on the terminal POPTR branch (3.25 for SG_Xa and 5.4 for SG_XIIb; Fig. 2). Surprisingly, SG_VIIa and SG_VIIb show a high branch-specific expansion rate in POPTR (4 and 3, respectively), which is not reflected in the global expansion rate in this species (Fig. 3). This is due to the fact that SG_VIIa and SG_VIIb went through strong reduction in Malpighiales (0.33) and fabids (0.5), respectively. Thus, the cumulative effect of successive reductions and expansions is not evident in the global expansion rate. These contrasted evolutionary dynamics can also be observed in MALDO. A global expansion of SG_IX was not detected because of the strong reduction in Amygdaloideae. To summarize, these data can be integrated into the species phylogeny to draw an image of the complex evolutionary dynamics of the LRR-RLK gene family through time (Fig. 4).

Different Patterns of Lineage-Specific Expansion in LRR-RLK Subgroups

Given the differences of LRR-RLK expansion rates between species, we wanted to identify cases of LSE (i.e. cases where a high duplication/retention rate is specific

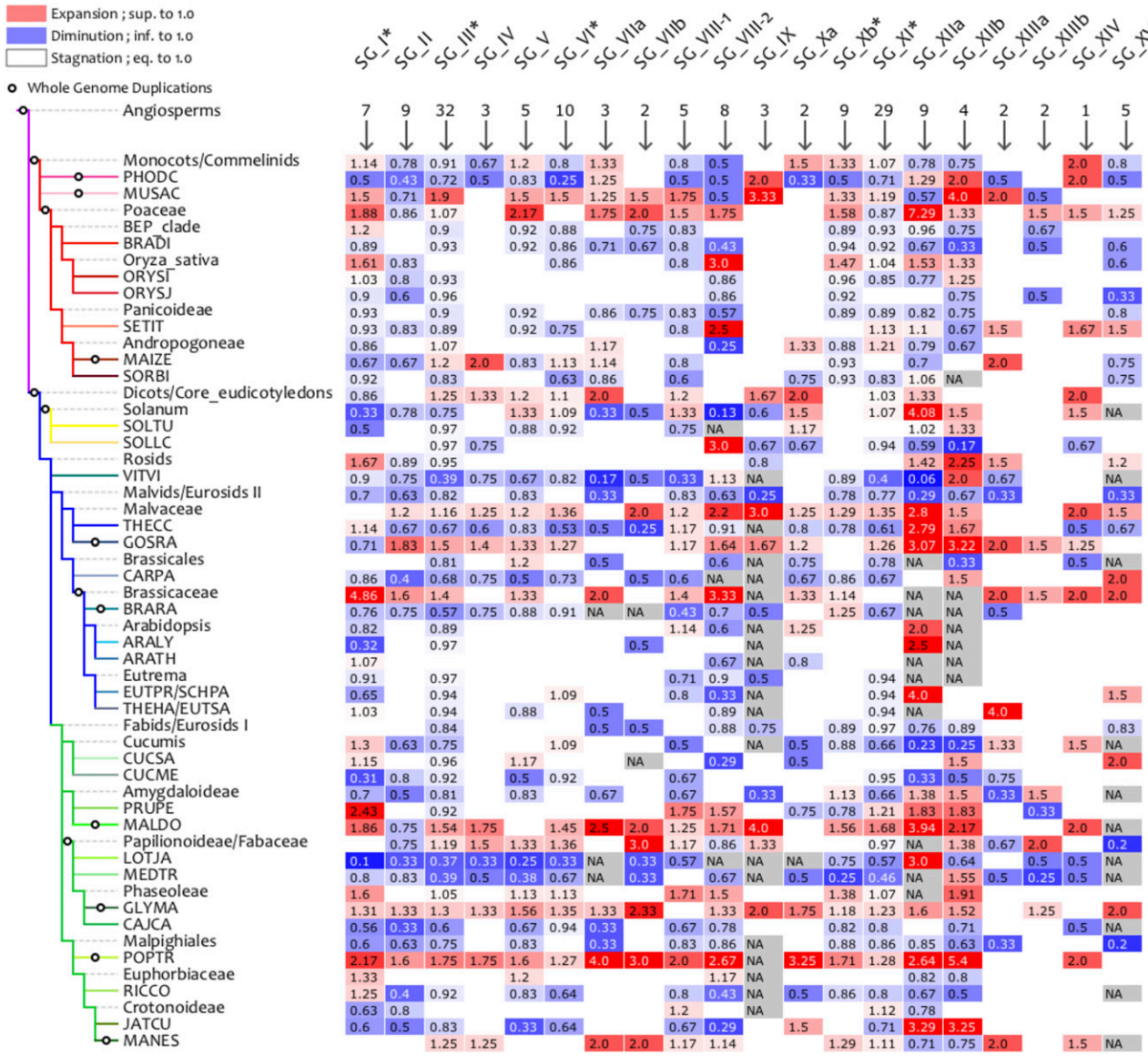

Figure 2. Branch-specific expansion/diminution of LRR-RLK genes for every SG on every branch in the phylogenetic tree. The tree on the left displays all the nodes and branches, and polyploidy events are marked with dots. Every line gives the expansion rate where the current (descendant) node is compared with the previous (ascendant) node. Red boxes indicate expansion, blue boxes indicate diminution, and blank boxes indicate stagnation. For example: SG_I* has the same number of copies in monocots compared with the ascendant node (angiosperms) indicated by a blank box. In PHODC, a diminution occurred compared with the ascendant node (monocots) indicated by a blue box. In MUSAC, an expansion occurred compared with the ascendant node (monocots) indicated by a red box, and so on.

to one species). Using a tree reconciliation approach (see "Materials and Methods"), we built a data set consisting of ultraparalog (UP; related only by duplication) clusters that represents the LSE events and a superortholog (SO; related only by speciation) reference gene set. We only considered clusters containing five or more sequences. After cleaning, our final data set comprised 75 UP and 189 SO clusters containing 796 and 1,970 sequences, respectively (Table IV). The median number of sequences in the UP clusters is not significantly different from the median number in the SO clusters (eight in both cases; Supplemental Fig. S1). For UP clusters, however, the alignments are significantly longer (Mann-Whitney test, $P < 0.001$), with a median of 3,237 bp compared with 2,841 bp for SO clusters. One possible explanation for this could be that UP clusters are more dynamic and might contain more LRRs. PRANK, the alignment algorithm we used, introduces gaps instead of aligning ambiguous sites and, therefore, produces longer alignments when sequences are divergent. However, this phenomenon does not

Table II. *Five-digit code for each species*

Species Name	Common Name	Five-Digit Code
Phoenix dactylifera	Date palm	PHODC
Musa acuminata	Banana	MUSAC
Brachypodium distachyon	Purple false brome	BRADI
Oryza sativa ssp. *japonica*	Asian rice	ORYSJ
Oryza sativa ssp. *indica*	Indian rice	ORYSI
Setaria italica	Foxtail millet	SETIT
Zea mays	Maize	MAIZE
Sorghum bicolor	Milo	SORBI
Solanum tuberosum	Potato	SOLTU
Solanum lycopersicum	Tomato	SOLLC
Vitis vinifera	Common grape vine	VITVI
Theobroma cacao	Cacao tree	THECC
Gossypium raimondii	Cotton progenitor	GOSRA
Carica papaya	Papaya	CARPA
Arabidopsis thaliana	Thale cress	ARATH
Arabidopsis lyrata	Outcrossing Arabidopsis relative	ARALY
Brassica rapa	Turnip	BRARA
Schrenkiella parvula	A saltwater cress	SCHPA
Eutrema salsugineum	A saltwater cress	EUTSA
Cucumis sativus	Cucumber	CUCSA
Cucumis melo	Melon	CUCME
Prunus persica	Peach	PRUPE
Malus × *domestica*	Apple	MALDO
Lotus japonicus		LOTJA
Medicago truncatula	Barrel medic	MEDTR
Glycine max	Soybean	GLYMA
Cajanus cajan	Pigeon pea	CAJCA
Populus trichocarpa	Black cottonwood	POPTR
Ricinus communis	Castor oil plant	RICCO
Jatropha curcas	Barbados nut	JATCU
Manihot esculenta	Cassava	MANES
Selaginella moellendorffii	A spikemoss	SELML
Physcomitrella patens	A moss	PHYPA

influence the outcome of further tests for positive selection using codeml (Yang, 2007).

We then wanted to determine which SGs are represented in the SO and UP data sets. Unsurprisingly, all SGs were present in SO clusters (Fig. 5). This was expected, as all SGs were already present in the LCAA and remained stable or expanded (except SG_XV*). In general, the frequency of SO clusters (and sequences) for each SG reflects the number of copies in the LCAA (Table I; Fig. 5). On the other hand, only 11 of the 20 SGs were represented in UP clusters (SG_I*, SG_III*, SG_VI*, SG_VIII-2, SG_IX, SG_Xa, SG_Xb*, SG_XI*, SG_XIIa, SG_XIIb, and SG_XIIIa), and these SGs harbor a total of 837 sequences. SG_I*, SG_VIII-2, SG_XIIa, and SG_XIIb are clearly overrepresented, which is in accordance with their expansion pattern. Other expanded SGs, however, have only a low number of UP clusters or, in the case of SG_IV, no UP clusters at all. Therefore, it seems that recently duplicated genes are more prone to be retained in some SGs.

Differences of Selective Constraint between Subgroups, Domains, and Amino Acids

To provide further insight into the LRR-RLK gene family evolution, we wanted to determine under which kind of selective pressures the LRR-RLK genes evolved. We focused on the data set described above (i.e. LSE and orthologous genes). We inferred the d_N/d_S ratio (or ω, i.e. the ratio of nonsynonymous to synonymous substitution rates) at codons of the alignments and branches of the phylogeny of the UP and SO clusters. An $\omega = 1$ indicates neutral evolution/relaxed constraint, an $\omega < 1$ indicates purifying selection, and an $\omega > 1$ can indicate positive selection. We used mapNH (Dutheil et al., 2012; Romiguier et al., 2012) to compute the ω for each branch. mapNH ran for 71 UP and 176 SO clusters containing 1,246 and 2,960 branches, respectively (Table IV). We first wanted to test for relaxation of selective constraint in UP and SO clusters and looked for branches with $\omega > 1$. We found 6.04% of UP branches but only 0.49% of SO branches to have an $\omega > 1$. The mean ω for branches with $\omega > 1$ is significantly larger in UP clusters (1.45) compared with SO clusters (1.13; $P = 0.004$). The same is true for branches with $\omega < 1$, where ω is significantly larger in UP clusters (0.48) compared with SO clusters (0.24; $P < 0.001$). Overall, the mean ω is significantly larger for branches from UP clusters (0.54) than for SO clusters (0.24; $P < 0.001$; Table IV; Supplemental Fig. S2).

We found 38 out of 75 UP clusters (50.67%) containing codons under positive selection (for details, see Supplemental Table S3) after manual curation but only six out of 186 SO clusters (3.23%). Additionally, codons under positive selection found in UP clusters are not distributed evenly over domains (Fig. 6). To account for the differences in domain size, a hit frequency (i.e. the number of sites under positive selection we found relative to all sites possible for each domain) was calculated (see "Materials and Methods"). The domain showing the highest hit frequency is the LRR domain, followed by the Cys pairs and their flanking regions (Fig. 6A). Hits in both domains are distributed over all SGs and species tested. The KD and its surrounding domains contain very few codons under positive selection. Domains classified as other combine domains important for the function of the LRR-RLK genes but vary between SGs. For example, SG_I* (Fig. 6B) contains a malectin domain. All hits classified as other here fall in the malectin-like domain of a POPTR SG_I* cluster.

Finally, we wanted to investigate whether some amino acids in the LRR are more frequently targeted by positive selection. The LRR typically contains 24 amino acids and sometimes islands between them (Fig. 6C). Four amino acids were predominantly subject to positive selection: 6, 8, 10, and 11, which all lie in the LRR-characteristic LXXLXLXX β-sheet/β-turn structure.

DISCUSSION

We studied the SG- and species-specific expansion dynamics in LRR-RLK genes from 31 angiosperm genomes in a phylogenetic framework. We also analyzed

Table III. *Estimated times of polyploidy events and corresponding references for Figure 4*

Event	Name	Reference	Age
			million years
1	Seed plant tetraploidy	Jiao et al. (2011)	350–330
2	Angiosperm tetraploidy	Jiao et al. (2011)	230–190
3	Monocot tetraploidy	Tang et al. (2010a)	130
4	Date palm WGD	D'Hont et al. (2012)	75–65 (?)
5	Banana gamma	D'Hont et al. (2012)	100
6	Banana beta	D'Hont et al. (2012)	65
7	Banana alpha	D'Hont et al. (2012)	65
8	Grass tetraploidy B (sigma)	D'Hont et al. (2012)	123–109
9	Grass tetraploidy (rho)	Paterson et al. (2004)	70
10	Maize tetraploidy	Schnable et al. (2011)	12–5
11	Eudicot hexaploidy (Arabidopsis gamma)	Jaillon et al. (2007); Cenci et al. (2010); Wang et al. (2012)	150–120
12	*Solanum* hexaploidy	Tomato Genome Consortium (2012)	91–52
13	Papiloniod tetraploidy	Pfeil et al. (2005)	55–54
14	Soybean tetraploidy	Pfeil et al. (2005)	15–13
15	Apple tetraploidy	Velasco et al. (2010); Verde et al. (2013)	45–30
16	Poplar tetraploidy	Tuskan et al. (2006)	65–60
17	Arabidopsis beta	Fawcett et al. (2009)	70–40
18	Arabidopsis alpha	Barker et al. (2009)	23
19	*Brassica* hexaploidy	Wang et al. (2011)	9–5
20	Cotton WGD	Wang et al. (2012)	20–13
21	Cassava WGD	Mühlhausen and Kollmar (2013)	? (after Crotonoideae split)

the lineage-specifically expanded genes in this family to determine to what extent positive selection occurred on them using a d_N/d_S-based test. We found differences in expansion patterns depending on SGs and species but only a few SGs that were subject to LSE. A significantly higher proportion of LSE LRR-RLK genes was affected by positive selection compared with single-copy genes, and the LRR domain (specifically four amino acids within this domain) was targeted by positive selection. In the following, we will discuss our findings in more detail.

Subgroup- and Species-Specific Expansions

We observed significant variations in the global expansion rates between LRR-RLK SGs. These are due to a complex history of expansion-retention-loss cycles that are specific to each SG. The phylogenetic approach allowed us to determine the relative importance of ancestral versus recent species-specific expansions for each SG and to characterize precisely the loss/retention dynamics during the evolutionary history of the studied species (summarized in Fig. 4). For example, SG_III* and SG_XI* had a high copy number of LRR-RLKs in the LCAA and kept a stable copy number over the last 150 million years. On the other hand, SG_I*, SG_XIIa, and SG_XIIb, which had a moderate copy number in the LCAA, keep expanding. Some functions have been described for genes of these SGs, mainly in Arabidopsis (Supplemental Table S4). For SG_III* and SG_XI*, mostly genes involved in development are described. The high numbers of ancestral genes in these two SGs combined with their size stability during angiosperm evolution may be interpreted as an early high level of diversification/specialization of these genes that are needed to orchestrate common developmental features. This hypothesis can be reinforced by the high number of superorthologous genes in these SGs. For SG_I* and SG_XIIa, on the other hand, mostly genes involved in responses to biotic stress are described at present. These observations confirm that different expansion/retention patterns appear to be related to gene function, although one has to keep in mind that functions have only been assigned to a few LRR-RLK genes. Three SGs (SG_IX, SG_Xa, and SG_XIV) expanded compared with their very low ancestral number (one to three), leading to a high total expansion rate. As it has been postulated that duplications are the raw material for adaptation (Nei and Rooney, 2005; Fischer et al., 2014), the evolution of those SGs was likely driven by adaptation, to varying degrees in different angiosperm species, depending on the environment they evolved in. The known functions are both related to responses to biotic or abiotic stress and development. Because so far our knowledge of LRR-RLK functions is limited and mostly restricted to Arabidopsis, further studies are needed to make more reliable statements on the link between function and expansion/retention dynamics in different SGs.

Next, we wanted to ascertain species-specific expansions of LRR-RLK genes and how they are related to the recent history of the species in our study. Whole-genome multiplication has been argued to be a major force in the diversification of angiosperms (Soltis et al., 2009; Soltis and Burleigh, 2009; Renny-Byfield and Wendel, 2014). All angiosperms share two ancient WGDs (Jiao et al., 2011). Likewise, all monocots share a WGD approximately 130 million years ago (Tang et al.,

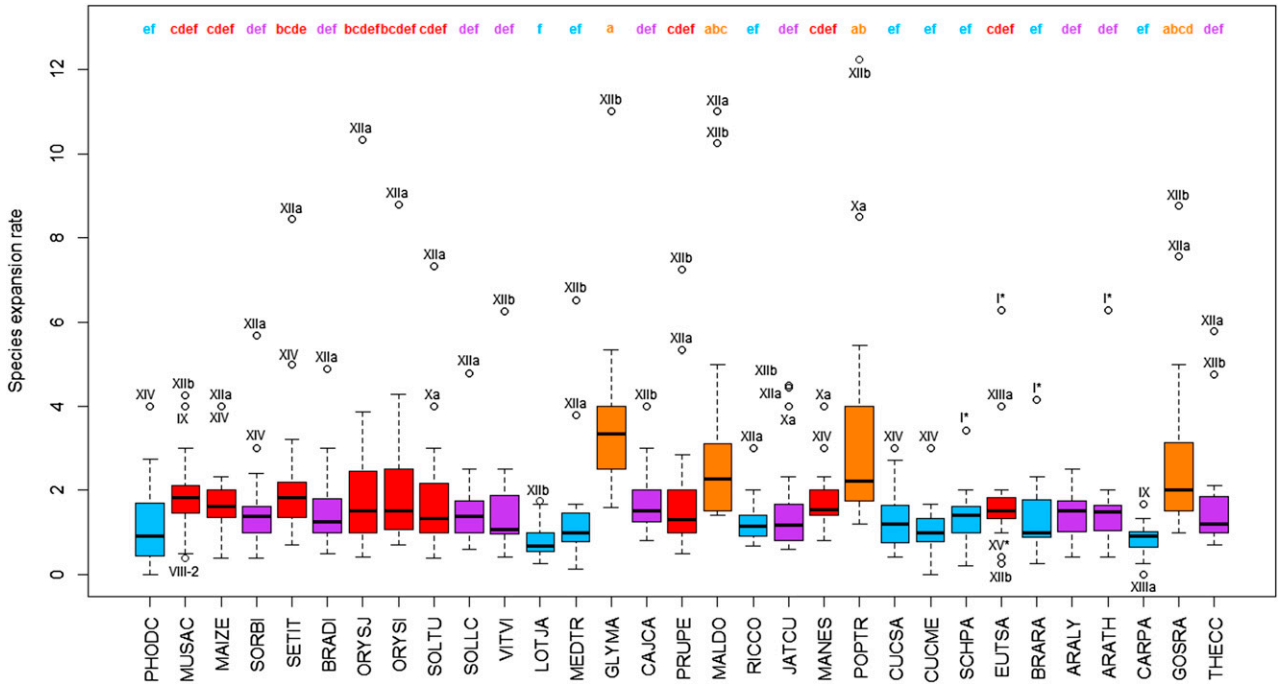

Figure 3. Global expansion rate in each species, which is the total number of genes in each species divided by the ancestral number (Table I). An ANOVA test showed that the expansion rate differs significantly between species ($P < 2e\text{-}16$). Therefore, we performed a TukeyHSD test to determine which species exactly show a significant difference between each other and grouped those species by significance level (a–e). Letters above each box plot indicate the TukeyHSD significance group (Supplemental Table S2). The significance groups are color coded according to the mean expansion rate: orange, greater than 2.25-fold expansion; red, 1.75- to 2.25-fold expansion; purple, 1.4- to 1.75-fold expansion; and blue, 0.8- to 1.4-fold expansion (i.e. no expansion). The outlier SGs are labeled for each species. For species identifiers, see Table II.

2010a), and most dicots (eudicots) share a WGT around the same time (Jaillon et al., 2007; Wang et al., 2012), but more recent WGDs and WGTs occurred in many angiosperm species (Fig. 4; Table III). The link between WGD/WGTs and the number of LRR-RLK genes is not straightforward. We found that in soybean (*Glycine max*), *Gossypium raimondii*, and apple (*Malus* × *domestica*), which were subject to relatively recent WGDs (15–13, 17–13, and 45–30 million years ago, respectively; Pfeil et al., 2005; Velasco et al., 2010; Wang et al., 2012), the number of LRR-RLK genes expanded more than 2-fold compared with the LCAA. These results are in accordance with what was already described for these species. Indeed, it was found that soybean contains a very large number of retained genes from this WGD (Cannon et al., 2015). Additionally, recent studies on large gene families in *G. raimondii* indicate that their copy number is driven either by retention after the last WGD (e.g. NAC transcription factors; Shang et al., 2013) or by a combination of segmental duplications (SDs) and tandem duplications (TDs; e.g. WRKY transcription factors; Dou et al., 2014). For apple (most recent WDG after the divergence for peach [*Prunus persica*] according to Verde et al. [2013]), a recent study on nucleotide-binding site LRR genes showed that they also stem mostly from SDs and TDs (Arya et al., 2014).

More contrasting results are observed in the Brassicaceae, where two WGDs occurred (Barker et al., 2009; Fawcett et al., 2009). Most SGs expand their number of genes on this ancestral branch, but the species belonging to this clade mostly retain or lose genes on average (Figs. 2 and 4). The only exception concerns *Eutrema salsugineum* (an Arabidopsis relative), which is the only species with a greater than 2-fold average expansion rate. The global expansion rate in *E. salsugineum* is mostly due to two SGs (SG_I* and SG_XIIIa). In the original genome study (Wu et al., 2012), the authors found that genes from the category response to stimulus (response to salt stress, osmotic stress, water deprivation, abscisic acid stimulus, and hypoxia) are significantly overrepresented in *E. salsugineum* compared with Arabidopsis. This overrepresentation is described as mostly caused by SDs and TDs (Wu et al., 2012), in accordance with what we observed in SG_XIIIa. This could be of functional importance to this halophile plant.

Finally, of all species analyzed here, maize (*Zea mays*) and *Brassica rapa* (and maybe *Manihot esculenta*) show the most recent cases of WGD/WGT (12–5 and 9–5 million years ago, respectively; Schnable et al., 2011; Wang et al., 2011), yet their expansion rates are moderate. This is further evidence for the dynamic nature of angiosperm genomes that has been discussed before

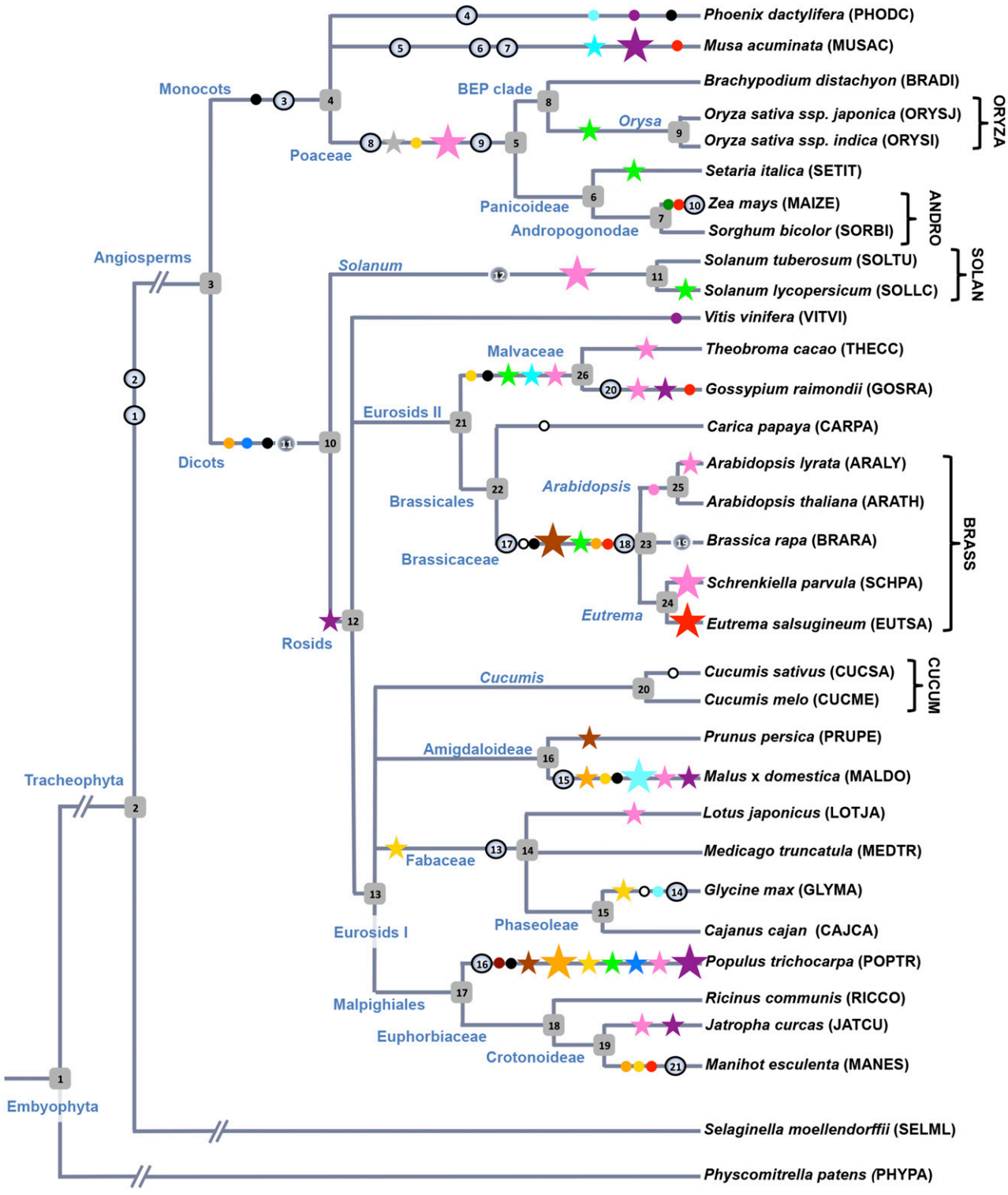

Figure 4. Phylogenetic tree of the 33 species studied here. Five-digit species identifiers are given in parentheses next to the species names. Species that diverged less than 15 million years ago were merged for the LSE analysis (see "Materials and Methods"): ANDRO, ORYZA, SOLAN, CUCUM, and BRASS. Polyploidy events and their estimated ages are indicated on the tree: circles on the branches represent WGD, and dark circles represent WGT. The numbers in the circles refer to details on the polyploidization events given in Table I. Species divergence and their estimated age are indicated by gray squares on the nodes. The numbers in the squares refer to details on the divergence times given in Supplemental Table S1. Dots and asterisks on the branches indicate SG expansions: dots, 2-fold; small asterisks, between 2- and 4-fold; and large asterisks, equal to or more than

Table IV. Details of the LSE and mapNH analyses for UP and SO clusters

Parameter	UP	SO
Total No. of clusters	75	189
Clusters for final mapNH analysis	71	176
Median cluster size (first; third Qu)	8 (6; 12)	8 (6; 14)
Minimum; maximum cluster size	5; 38	5; 25
Median alignment length (first; third Qu)	3,237 (2,952; 3,574)	2,841 (2,034; 3,192)
Minimum; maximum alignment length	1,749; 8,691	861; 6,216
Branches analyzed/total No. of branches	1,193/1,246	2,860/2,960
Clusters with branches $\omega > 1$ (%)	25 (35.21)	10 (5.68)
Branches with $\omega < 1$ (%)	1,121 (93.96)	2,846 (99.51)
Mean ω for less than one branch ± SD	0.48 ± 0.17	0.24 ± 0.12
Branches with $\omega > 1.0$ (%)	72 (6.04)	14 (0.49)
Mean ω for more than one branch ± SD	1.45 ± 0.51	1.13 ± 0.14
Mean ω ± SD	0.54 ± 0.31	0.24 ± 0.13

(Leitch and Leitch, 2012; Fischer et al., 2014). After a WGD event, genomes tend to return to the diploid (or previous) state by losing redundant duplicated genes (fractionation process), although the gene loss is biased (Bowers et al., 2003; Schnable et al., 2009). Which genes are lost or retained depends strongly on their function (De Smet et al., 2013). However, it has been shown that genes involved in stress responses are mostly created by TD rather than WGD (Hanada et al., 2008). Indeed, it was hypothesized before that RLK genes involved in stress responses mostly duplicate by TD (Shiu et al., 2004). Here, we provide a detailed representation of expansion-retention-loss dynamics of the whole LRR-RLK gene family in 31 angiosperm species (Fig. 4). Each new genome sequenced will improve the accuracy of the expansion-retention-loss event predictions and will help in identifying new elements that can be useful for future functional analysis and/or linked to adaptive traits.

Studying Selection Pressures in a Large and Dynamic Gene Family

As described above, the composition of LRR-RLKs in each of the 31 studied angiosperm species results from a complex dynamic of species- and SG-specific expansion/loss events. To further investigate the potential role of this family in plant adaptation, we analyzed the selective pressures to which the LRR-RLKs were subject. Such an analysis cannot be considered for the phylogeny of the entire gene family because of the high number of sequences and the high sequence divergence (the phylogeny on which we divided the SGs was inferred on the conserved KD only). We then chose to focus on two specific cases: (1) LSE as a specific case of duplication/retention, and (2) a subset of strictly orthologous genes. Indeed, LSE has been shown to fuel adaptation in angiosperms (Fischer et al., 2014), and we wanted to test the prevalence of this mode of duplication in our large data set. Therefore, we evaluated the extent to which LRR-RLK genes were subject to LSE and how positive selection acted on those genes. As a reference, we chose the strictly orthologous subset. This approach allows the interpretation of LSE evolution compared with the general LRR-RLK selective background (Fischer et al., 2014).

The power of this phylogenetic approach relies on the number of species analyzed, and we profit from an ever-increasing number of sequenced plant genomes. Another important requirement for this approach is the quality of sequencing and annotation, especially for a large gene family, as sequencing errors and misannotations can lead to false positives when testing for positive selection (Han et al., 2013). We profit from a recently developed pipeline designed to automatically perform different steps of the analysis (Fischer et al., 2014). This allowed us to quickly incorporate sequenced genomes of choice, and future studies can easily expand this analysis as new reliable data become available. Finally, we set great value on manually verifying the data throughout the process, from the identification of the LRR-RLKs to the inference of positive selection. Although this is tedious work for such a large data set, it is important nevertheless. As we recently showed, approximately 50% automatically reported instances of positive selection turned out to be false positives after manual curation (Fischer et al., 2014).

We found that all SGs are represented in the single-copy reference set, with an overrepresentation of SG_III* and SG_XI*. This is in accordance with the fact that these two SGs had the highest number of copies in the genome of the LCAA and did not expand significantly since (see above). In general, the frequency of clusters from the single-copy gene set (and sequences) for each SG reflects the number of copies in the LCAA (Table I; Fig. 5). On the other hand, only 11 of the 20 SGs

Figure 4. (*Continued.*)
4-fold. SGs are as follows: SG I* (brown), SG_IV (dark green), SG_V (gray), SG_VIIa (orange), SG_VIIb (yellow), SG_VIII-1 (dark brown), SG_VIII-2 (green), SG_IX (light blue), SG_Xa (dark blue), SG_XIIa (pink), SG_XIIb (purple), SG_XIIIa (red), SG_XIV (black), and SG_XV* (white). The asterisks and dots do not indicate the exact age.

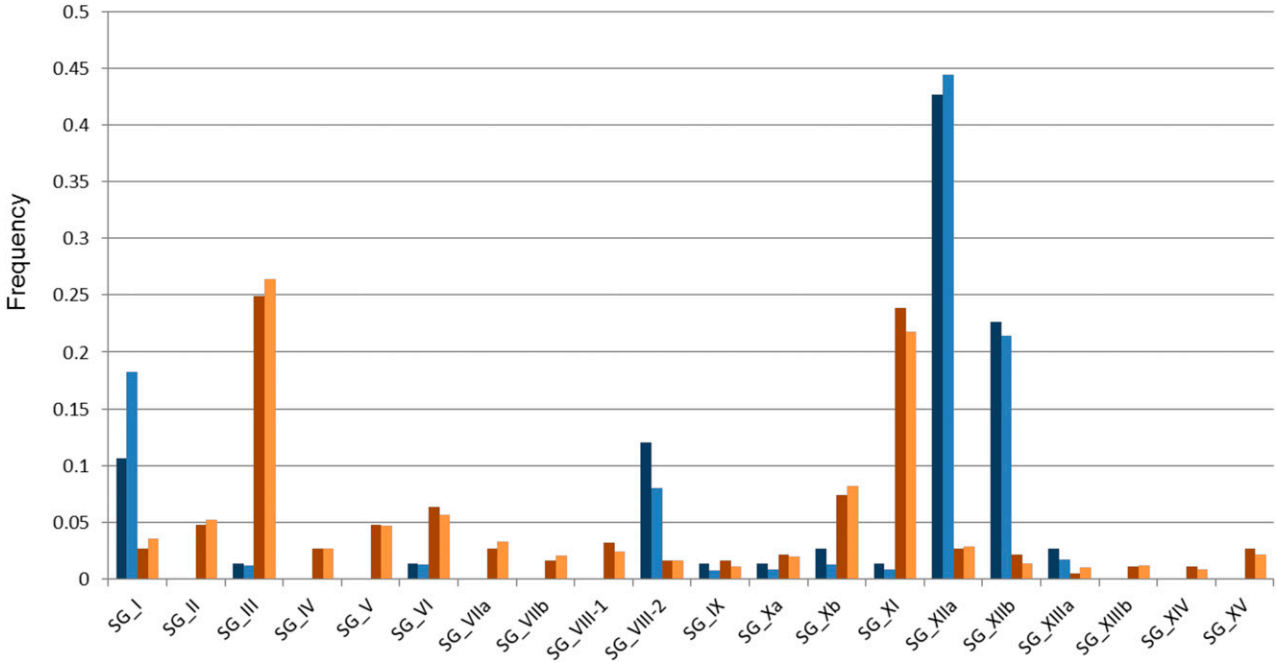

Figure 5. Distribution of UP and SO clusters and sequences across all SGs. The frequency of all extracted UP (dark blue) and SO (dark orange) clusters for each SG, and the frequency of all extracted UP (light blue) and SO sequences (light orange) for each SG, are shown.

were represented in the LSE data set. This is mainly because the majority of expansions are rather old in these SGs, whereas they happened relatively recently in SG_I*, SG_VIII-2, SG_XIIa, and SG_XIIb (see above). Fourteen species (or clades) are represented in the LSE data set: MUSAC (two UP clusters), SETIT (one), ORYZA (10), VITVI (three), SOLAN (six), MEDTR (three), GLYMA (two), PRUPE (six), MALDO (11), POPTR (eight), BRASS (11), GOSRA (five), THECC (two), and PHYPA (five). Again, not every species is affected to the same extent, but this does not necessarily reflect recent WGD/WGT. Additionally, LSE can also arise from SD and TD, the frequency of which is not uniform within or between genomes. Our results indicate that different species are more likely to retain recently duplicated genes than others. This, in turn, might reflect on their recent evolution or domestication, which should be examined in more detail in future studies.

When focusing on the study of selective pressures, we first looked at ω at the branches of the LSE and single-copy gene clusters and found that selective constraint was relaxed in the LSE data set. This outcome was expected, as it was shown previously that LSE genes evolve more relaxed constraint than single-copy genes in angiosperms (Fischer et al., 2014). This study, however, looked at whole angiosperm genomes, but a similar pattern has already been demonstrated in other large gene families (Johnson and Thomas, 2007; Xue et al., 2012; Yang et al., 2013a, 2013b) and in LRR-RLK genes in particular (Tang et al., 2010b). Previous studies on that subject only had a limited data set (four angiosperm species; Tang et al., 2010b). Here, we demonstrate that this is still true when a larger and more representative sample of angiosperms is considered.

Next, we wanted to identify codons that evolved under positive selection in the LSE and the single-copy data sets. A recent study on gene families in the whole genomes of 10 angiosperms found that 5.4% of LSE genes contained codons showing positive selection footprints (Fischer et al., 2014). Here, we ask if and to what extent this is also true for the large and dynamic LRR-RLK gene family. We discovered that for LSE LRR-RLK genes, the rate of codons under selection is almost 10-fold higher (50.67%) than the genome average. In addition, we found more than 3% of single-copy genes containing codons under selection, whereas Fischer et al. (2014) described no case of positive selection at the single-copy gene clusters in their study. Together with the high rate of branches with $\omega > 1$ in LSE gene clusters (6.04%, compared with 0.49% for single-copy genes), this indicates that LRR-RLK genes are more prone to evolve under positive selection than the average for angiosperm gene families. As might be expected, all UP clusters with codons under positive selection come from the four overrepresented SGs: SG_I* (one UP cluster), SG_VIII-2 (three), SG_XIIa (24), and SG_XIIb (10). The single-copy gene clusters with codons under selection come from six SGs: SG_III, SG_VIIa, SG_Xa, SG_Xb, SG_XIIa, and SG_XIIb. Therefore, recent expansion and retention affect only a few SGs, but in those SGs positive selection plays an important role. For SG_XIIa, positive selection has been inferred previously for genes involved in environmental interactions: *Xa21*, which confers resistance to the bacterial blight disease, was found to have

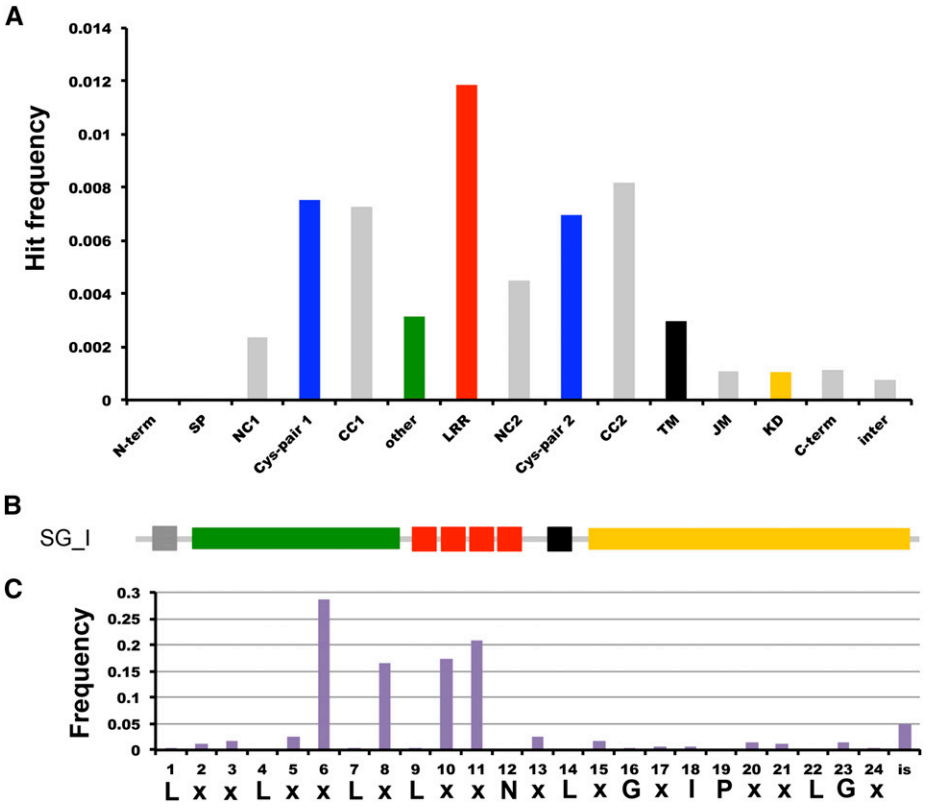

Figure 6. A, Hit frequency (i.e. frequency of codons under selection versus the total number of sites) for each domain of the LRR-RLK genes. The absence/presence and size of the domains vary between SGs, for details, see text. N-term, N-Terminal end; SP (dark gray), signal peptide; NC1, N-terminal end of Cys-pair 1; Cys-pair 1 (blue), first Cys pair; CC1, C-terminal end of Cys-pair 1; other (green), other domains; NC2, N-terminal end of Cys-pair 2; Cys-pair 2 (blue), second Cys pair; CC2, C-terminal end of Cys-pair 2; TM (black), transmembrane domain; JM, juxtamembrane domain; C-term, C-terminal end; inter, other interdomain regions. B, Schematic structure of the LRR-RLK genes, here with SG_I* gene structure as an example. C, Frequency of amino acids in the LRR domain under positive selection. L, Leu; x, variable; N, Asn; G, Gly; I, Ile; P, Pro; is, island between LRRs.

evolved under positive selection in rice (Wang et al., 1998; Tan et al., 2011); and FLS2, involved also in responses to biotic stress, shows a signature of rapid fixation of an adaptive allele in Arabidopsis (Vetter et al., 2012). Future studies on smaller subsets of SGs will surely cast further light on selection patterns in LRR-RLK genes. Only 11 species (or clades) are represented in the LSE data set with codons under positive selection: SETIT (one UP cluster), ORYZA (two), SOLAN (four), MEDTR (two), GLYMA (two), PRUPE (three), MALDO (eight), POPTR (seven), BRASS (two), GOSRA (five), and THECC (two). Not every species is affected to the same extent by positive selection, and again, future studies might bring more details concerning the evolutionary history of specific species and SGs to light.

In addition, we found that not every domain of the LRR-RLK genes was similarly affected by positive selection. Most codons under selection fall in the LRR domain. This outcome might be expected, as LRRs are very dynamic and plasticity in this region provides plants with a broad tool set to face environmental challenges and, therefore, undergoes positive selection frequently (Zhang et al., 2006; Tang et al., 2010b). Only very few codons under positive selection were found in the KD and its surrounding regions. This result is consistent with the fact that the KD is very conserved among species and SGs and evolved mostly under purifying selection (Shiu et al., 2004; Tang et al., 2010b). A more surprising result was the identification of a significant number of positively selected sites in the malectin-like domain of a poplar SG_I* cluster. So far, the function of extracellular malectin-like domains of RLKs is not well understood (Lindner et al., 2012). However, a malectin-like domain-containing SG_I* LRR-RLK has been described to confer susceptibility to a downy mildew pathogen in Arabidopsis and to have similarities to symbiosis RLKs, which are important for the regulation of bacterial symbiont accommodation (Markmann et al., 2008; Hok et al., 2011). Therefore, our results suggest that it could be interesting to further investigate the function and evolutionary history of this SG_I* domain, particularly in poplar. Another unexpected finding was the frequent occurrence of positive selection at the Cys pairs and flanking regions that are involved in folding and/or the binding to other proteins. To what extent the function of LRR-RLKs is affected by mutations in the Cys pair regions depends on the function of the gene (Song et al., 2010; Sun et al., 2012), and it would be interesting to study this in more detail in the future.

Finally, we took a closer look at the amino acids in the LRR primarily affected by positive selection. Only four, out of the 24 amino acids an LRR typically contains, were predominantly and strongly subject to positive selection. These variable amino acids lie in the unconserved part of the LRR-characteristic LXXLXLXX β-sheet/β-turn structure, which is involved in protein-protein interactions (Jones and Jones, 1997; Enkhbayar et al., 2004). Specifically, solvent-exposed residues were targeted by positive selection (Parniske et al., 1997;

Wang et al., 1998). Further investigation of the functional consequences of these nucleotide variations need to be done to confirm their adaptive potential, but our findings align very well with the current understanding of LRR ligand binding. Taken together, our results could be very useful for further functional investigations of LRR-RLK genes in different species.

CONCLUSION

We studied LRR-RLK genes from 33 land plant species to investigate SG- and species-specific expansion of these genes, the extent to which they were subject to LSE, and the role that positive selection played in the evolution of this large gene family. We described that some SGs are more prone to expansion/retention than others and that the expansions occurred at different times in the evolution of LRR-RLK genes. This fine-scale analysis of the dynamic allowed us to identify branches and species for which a higher than average retention rate could indicate a potential adaptive event for some SGs. We also described that only a few SGs show patterns of recent LSE and that, at those genes, selective constraint is relaxed. More than 50% of the LSE genes contain codons that show evidence for positive selection, which is almost 10-fold the frequency described previously for gene families in angiosperms (Fischer et al., 2014). Finally, we found that, across the LRR-RLK genes, the LRR domain and specifically four amino acids responsible for ligand interaction are most frequently subject to selection.

MATERIALS AND METHODS

Studied Genomes

We analyzed 31 angiosperm genomes (eight monocot [sub]species and 23 dicot species; Table II): *Phoenix dactylifera* (Al-Dous et al., 2011), *Musa acuminata* (D'Hont et al., 2012), *Oryza sativa* ssp. *japonica* (International Rice Genome Sequencing Project, 2005), *Oryza sativa* ssp. *indica* (Yu et al., 2002), *Brachypodium distachyon* (International Brachypodium Initiative, 2010), *Zea mays* (Schnable et al., 2009), *Sorghum bicolor* (Paterson et al., 2009), *Setaria italica* (Zhang et al., 2012), *Solanum tuberosum* (Xu et al., 2011), *Solanum lycopersicum* (Tomato Genome Consortium, 2012), *Vitis vinifera* (Jaillon et al., 2007), *Lotus japonicus* (Sato et al., 2008), *Cajanus cajan* (Varshney et al., 2012), *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), *Arabidopsis lyrata* (Hu et al., 2011), *Schrenkiella parvula* (a synonym is *Eutrema parvula*; we used the nomenclature from Oh et al. [2014]; Dassanayake et al., 2011), *Eutrema salsugineum* (a synonym is *Thellungiella halophila*; we chose the nomenclature according to Phytozome [http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Esalsugineum]; Wu et al., 2012), *Brassica rapa* (Wang et al., 2011), *Populus trichocarpa* (Tuskan et al., 2006), *Glycine max* (Schmutz et al., 2010), *Medicago truncatula* (Young et al., 2011), *Prunus persica* (Ahmad et al., 2011), *Malus* × *domestica* (Velasco et al., 2010), *Ricinus communis* (Chan et al., 2010), *Jatropha curcas* (Sato et al., 2011), *Manihot esculenta* (Prochnik et al., 2012), *Cucumis sativus* (Huang et al., 2009), *Cucumis melo* (Garcia-Mas et al., 2012), *Carica papaya* (Ming et al., 2008), *Gossypium raimondii* (Wang et al., 2012), and *Theobroma cacao* (Argout et al., 2011). We also extracted LRR-RLKs from the moss *Physcomitrella patens* (Rensing et al., 2008) and the spikemoss *Selaginella moellendorffii* (Banks et al., 2011). Throughout this article, we refer to the species using five-digit identifiers, which can be found in Table II. Altogether, we analyzed 33 genomes from 39 proteomes (we used several annotation versions of the Arabidopsis and rice genomes). Details on which genome versions we used can be found in Supplemental Table S5. The phylogeny of those species is provided in Figure 4.

LRR-RLK Extraction, Clustering, Phylogeny, and Identification of Gain/Loss Events

We used the hmmsearch program (Eddy, 2009) to extract peptide sequences containing both intact (i.e. nondegenerated) LRR(s) and a KD from the proteomes as described previously (Diévart et al., 2011). We classified SGs using the KD by a global phylogenetic analysis (the tree can be found at http://phylogeny.southgreen.fr/kinase/index.php; Global Analysis). First, sequences were aligned using MAFFT (Katoh et al., 2005) with a progressive strategy. Second, the alignments were cleaned using trimAl (Capella-Gutiérrez et al., 2009) with settings to remove every site with more than 20% gaps or with a similarity score lower than 0.001. Third, a similarity matrix was computed by ProtDist (Felsenstein, 1993) using a JTT model. Fourth, a global distance phylogeny was inferred using FastME (Desper and Gascuel, 2006) with default settings and SPR movements to optimize the tree topology. Fifth, SGs were defined manually in the global phylogeny using the Arabidopsis genes as reference, which led us to 20 SGs in contrast to the 15 described previously (Shiu et al., 2004; Lehti-Shiu et al., 2009).

More accurate phylogenies were then inferred for each of the 20 SGs. The KDs of the sequences attributed to each SG were realigned using MAFFT with an iterative strategy (maximum of 100 iterations). Alignments were cleaned using trimAl with settings to only remove sites with more than 80% gaps. Then, maximum likelihood phylogenies were inferred by PhyML 3.0 (Guindon et al., 2010) using an LG+gamma model and the best of NNI and SPR topology optimization. Statistical branch support was computed using the aLRT/SH-like strategy (Guindon et al., 2010). This left us with 20 phylogenies, one for each SG (all phylogenies are available at http://phylogeny.southgreen.fr/kinase/index.php; SG_I–SG_XV).

Each of the 20 phylogenetic trees has been reconciled with the species tree using RAP-Green (Dufayard et al., 2005; https://github.com/SouthGreenPlatform/rap-green). By comparing the gene tree with the species tree, this analysis allows us to root phylogenetic trees and to infer duplication and loss events (Dufayard et al., 2005). We tested this approach of rooting (by minimizing the number of inferred duplications and losses) and compared it with rooting with outgroups (data not shown). The two methods provided very close root locations that did not change the overall conclusions. Using this RAP-Green tree reconciliation approach (for parameters, the maximum support for reduction is 0.95), we inferred the number of duplications and losses at each node of the species tree. Briefly, each duplication and loss increases and decreases, respectively, by one the number of copies in the common ancestor of the taxonomic group analyzed.

We determined the global SG- and species-specific expansion rates by computing the number of LRR-RLK genes in one SG divided by the ancestral number and the number of LRR-RLK genes in one species divided by the ancestral number, respectively. An ANOVA showed that the expansion rate differed significantly between the SGs/species ($P < 2e\text{-}16$ in both cases). We used the TukeyHSD test of the agricolae package (http://cran.r-project.org/web/packages/agricolae/index.html) in R (R Development Core Team, 2012) to further explore which groups of SGs/species differ from each other. This test compares the range of sample means and defines an honestly significance difference value, which is the minimum distance between groups to be considered statistically significant. In short, TukeyHSD is a posthoc test that groups subsets by significance levels after ANOVA showed significant differences between subsets.

LSE Data Set and Testing for Positive Selection

Testing for adaptation can be done by comparing positive (Darwinian) selection footprints in lineages with recently and specifically duplicated genes to reference lineages containing only single-copy genes. One way to infer positive selection is by analyzing nucleotide substitution data at the codon level in a phylogenetic framework. As nucleotide substitutions can be either nonsynonymous (i.e. protein changing, thereby potentially impacting the fitness) or synonymous (i.e. not protein changing, thereby theoretically without consequences for the fitness; Lawrie et al., 2013), the nonsynonymous/synonymous substitution rate ratio, denoted as d_N/d_S or ω, can be used to infer the direction and strength of natural selection. An $\omega < 1$ indicates purifying selection, and the closer ω is to 0, the stronger purifying selection is acting. Under neutral evolution, $\omega = 1$. An $\omega > 1$ indicates that positive selection is acting.

We identified UP clusters (related only by duplication) using a tree reconciliation approach (Dufayard et al., 2005). Those represent our LSE gene set. As a single-copy gene reference, we chose an SO gene set (related only by speciation). We chose clusters with a minimum of five sequences. To address the

question of whether positive selection is more frequent after LSE events, we compared the results obtained on UPs with those obtained on SO gene sets. Species that diverged less than 15 million years ago were merged for the LSE detection (Fig. 4) in order not to overly reduce the UP data set and to not induce bias due to very recent speciation events: ANDRO (ZEAMA and SORBI), ORYZA (ORYSJ and ORYSI), SOLAN (SOLLC and SOLTU), CUCUM (CUCSA and CUCME), and BRASS (ARATH, ARALY, BRARA, SCHPA, and EUTSA). We then applied the pipeline developed by Fischer et al. (2014) to the extracted UP and SO clusters. In short, the pipeline consists of the following steps. (1) The clusters were aligned using $PRANK_{+F}$ with codon option (Löytynoja and Goldman, 2005). The alignments were cleaned by GUIDANCE (Penn et al., 2010) with the default sequence quality cutoff and a column cutoff of 0.97 to remove problematic sequences and unreliable sites from the alignments. We used PRANK and GUIDANCE here because previous benchmarks (Fletcher and Yang, 2010; Jordan and Goldman, 2012) showed that these programs lead to a minimum of false positives when inferring positive selection using codeml. The cleaned alignments can be retrieved at http://phylogeny.southgreen.fr/kinase/alignments.php (manually curated alignments for positive selection analysis). (2) We relied on the EggLib package (De Mita and Siol, 2012) to infer the maximum likelihood phylogeny at the nucleotide level for every alignment using PhyML 3.0 (Guindon et al., 2010) under the GTR substitution model. (3) We ran the codeml site model implemented in the PAML4 software (Yang, 2007) to infer positive selection on codons under several substitution models. In clusters identified to have evolved under positive selection, Bayes empirical Bayes was used to calculate the posterior probabilities at each codon and detect those under positive selection (i.e. those with a posterior probability of $\omega > 1$ strictly above 95%). All alignments detected to be under positive selection at the codon level were curated manually for potential alignment errors. Details of all codons showing a signal of positive selection using codeml can be found in Supplemental Table S3. (4) We used mapNH (Dutheil et al., 2012; Romiguier et al., 2012) to infer ω at the branch level.

In order to analyze the distribution of positively selected sites among domains, we calculated a hit frequency that computes the number of sites under positive selection found in each domain relative to all sites possible. All possible sites for each domain were calculated as follows. First, we extracted the size of each domain of every SG. If SGs were subdivided further, we took the average size of each domain. Second, we multiplied the size of each domain by the number of UP clusters we found for each SG. For example: the LRR of SG_I* contains an average of 77 sites, and we found eight UP clusters for SG_I*. Therefore, the total number of possible LRR sites for SG_I* is $77 \times 8 = 616$ sites. Third, we added up the sites for each domain for all SGs.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Summary of UP and SO cluster size and length.

Supplemental Figure S2. ω distribution of branches of UP and SO clusters.

Supplemental Table S1. Estimated divergence times and corresponding references for Figure 4.

Supplemental Table S2. Results of the TukeyHSD test.

Supplemental Table S3. Details of all codons showing a signal of positive selection using codeml.

Supplemental Table S4. Arabidopsis LRR-RLK gene classification according to The Arabidopsis Information Resource.

Supplemental Table S5. List of genomes used here, with name, link, and version of the genome fasta file.

ACKNOWLEDGMENTS

We thank the reviewers for their helpful comments.

Received September 16, 2015; accepted January 14, 2016; published January 15, 2016.

LITERATURE CITED

Ahmad R, Parfitt DE, Fass J, Ogundiwin E, Dhingra A, Gradziel TM, Lin D, Joshi NA, Martinez-Garcia PJ, Crisosto CH (2011) Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. BMC Genomics **12:** 569

Albert M, Jehle AK, Mueller K, Eisele C, Lipschis M, Felix G (2010) *Arabidopsis thaliana* pattern recognition receptors for bacterial elongation factor Tu and flagellin can be combined to form functional chimeric receptors. J Biol Chem **285:** 19035–19042

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, et al (2011) *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat Biotechnol **29:** 521–527

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, et al (2011) The genome of *Theobroma cacao*. Nat Genet **43:** 101–108

Arya P, Kumar G, Acharya V, Singh AK (2014) Genome-wide identification and expression analysis of NBS-encoding genes in *Malus* × *domestica* and expansion of NBS genes family in Rosaceae. PLoS ONE **9:** e107987

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, et al (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science **332:** 960–963

Barker MS, Vogel H, Schranz ME (2009) Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. Genome Biol Evol **1:** 391–399

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B, Stewart CN, Rolf M, et al (2015) Multiple polyploidy events in the early radiation of nodulating and non-nodulating legumes. Mol Biol Evol **32:** 193–210

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25:** 1972–1973

Castells E, Casacuberta JM (2007) Signalling through kinase-defective domains: the prevalence of atypical receptor-like kinases in plants. J Exp Bot **58:** 3503–3511

Cenci A, Combes MC, Lashermes P (2010) Comparative sequence analyses indicate that *Coffea* (asterids) and *Vitis* (rosids) derive from the same paleo-hexaploid ancestral genome. Mol Genet Genomics **283:** 493–501

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol **28:** 951–956

Chevalier D, Batoux M, Fulton L, Pfister K, Yadav RK, Schellenberg M, Schneitz K (2005) *STRUBBELIG* defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis*. Proc Natl Acad Sci USA **102:** 9074–9079

Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, et al (2011) The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet **43:** 913–918

De Mita S, Siol M (2012) EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet **13:** 27

De Smet I, Voss U, Jürgens G, Beeckman T (2009) Receptor-like kinases shape the plant. Nat Cell Biol **11:** 1166–1173

De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci USA **110:** 2898–2903

Desper R, Gascuel O (2006) Getting a tree fast: neighbor joining, FastME, and distance-based methods. Curr Protoc Bioinformatics **Chapter 6:** 6.3.1–6.3.28

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature **488:** 213–217

Diévart A, Gilbert N, Droc G, Attard A, Gourgues M, Guiderdoni E, Périn C (2011) Leucine-rich repeat receptor kinases are sporadically distributed in eukaryotic genomes. BMC Evol Biol **11:** 367

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet **11:** 539–548

Dou L, Zhang X, Pang C, Song M, Wei H, Fan S, Yu S (2014) Genome-wide analysis of the WRKY gene family in cotton. Mol Genet Genomics **289:** 1103–1121

Dufayard JF, Duret L, Penel S, Gouy M, Rechenmann F, Perrière G (2005) Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. Bioinformatics **21:** 2596–2603

Dutheil JY, Galtier N, Romiguier J, Douzery EJ, Ranwez V, Boussau B (2012) Efficient selection of branch-specific models of sequence evolution. Mol Biol Evol **29:** 1861–1874

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform **23:** 205–211

Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, Matsushima N (2004) Structural principles of leucine-rich repeat (LRR) proteins. Proteins Struct Funct Bioinf **54:** 394–403

Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci USA **106:** 5737–5742

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Washington, Seattle

Fischer I, Dainat J, Ranwez V, Glémin S, Dufayard JF, Chantret N (2014) Impact of recurrent gene duplication on adaptation of plant genomes. BMC Plant Biol **14:** 151

Fletcher W, Yang Z (2010) The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. Mol Biol Evol **27:** 2257–2267

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, et al (2012) The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci USA **109:** 11872–11877

Gish LA, Clark SE (2011) The RLK/Pelle family of kinases. Plant J **66:** 117–127

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:** 307–321

Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered **100:** 605–617

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol **30:** 1987–1997

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol **148:** 993–1003

Hok S, Danchin EGJ, Allasia V, Panabières F, Attard A, Keller H (2011) An *Arabidopsis* (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant Cell Environ **34:** 1944–1957

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet **43:** 476–481

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet **41:** 1275–1281

Innan H (2009) Population genetic models of duplicated genes. Genetica **137:** 19–37

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet **11:** 97–108

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436:** 793–800

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449:** 463–467

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature **473:** 97–100

Johnson DA, Thomas MA (2007) The monosaccharide transporter gene family in *Arabidopsis* and rice: a history of duplications, adaptive evolution, and functional divergence. Mol Biol Evol **24:** 2412–2423

Jones DA, Jones JDG (1997) The role of leucine-rich repeat proteins in plant defences. Adv Bot Res **24:** 90–167

Jordan G, Goldman N (2012) The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol Biol Evol **29:** 1125–1139

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res **33:** 511–518

Langham RJ, Walsh J, Dunn M, Ko C, Goff SA, Freeling M (2004) Genomic duplication, fractionation and the origin of regulatory novelty. Genetics **166:** 935–945

Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong purifying selection at synonymous sites in *D. melanogaster*. PLoS Genet **9:** e1003527

Lehti-Shiu M, Zou C, Shiu SH (2012) Origin, diversity, expansion history, and functional evolution of the plant Receptor-Like Kinase/*Pelle* family. *In* F Tax, B Kemmerling, eds, Receptor-Like Kinases in Plants, **Vol 13**. Springer, Berlin, pp 1–22

Lehti-Shiu MD, Zou C, Hanada K, Shiu SH (2009) Evolutionary history and stress regulation of plant *Receptor-Like Kinase/Pelle* genes. Plant Physiol **150:** 12–26

Leitch AR, Leitch IJ (2012) Ecological and genetic factors linked to contrasting genome dynamics in seed plants. New Phytol **194:** 629–646

Lindner H, Müller LM, Boisson-Dernier A, Grossniklaus U (2012) CrRLK1L receptor-like kinases: not just another brick in the wall. Curr Opin Plant Biol **15:** 659–669

Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci USA **102:** 10557–10562

Markmann K, Giczey G, Parniske M (2008) Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. PLoS Biol **6:** e68

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, et al (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature **452:** 991–996

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol **8:** 122–128

Mühlhausen S, Kollmar M (2013) Whole genome duplication events in plant evolution reconstructed and predicted using myosin motor proteins. BMC Evol Biol **13:** 202

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet **39:** 121–152

Oh DH, Hong H, Lee SY, Yun DJ, Bohnert HJ, Dassanayake M (2014) Genome structures and transcriptomes signify niche adaptation for the multiple-ion-tolerant extremophyte *Schrenkiella parvula*. Plant Physiol **164:** 2123–2138

Oh MH, Wang X, Kota U, Goshe MB, Clouse SD, Huber SC (2009) Tyrosine phosphorylation of the BRI1 receptor kinase emerges as a component of brassinosteroid signaling in *Arabidopsis*. Proc Natl Acad Sci USA **106:** 658–663

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BBH, Jones JDG (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell **91:** 821–832

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551–556

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA **101:** 9903–9908

Penn O, Privman E, Landan G, Graur D, Pupko T (2010) An alignment confidence score capturing robustness to guide tree uncertainty. Mol Biol Evol **27:** 1759–1767

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol **54:** 441–454

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, et al (2012) The cassava genome: current progress, future directions. Trop Plant Biol **5:** 88–94

R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Bot **101:** 1711–1725

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, et al (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science **319:** 64–69

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLOS Comput Biol **2:** e115

Romiguier J, Figuet E, Galtier N, Douzery EJ, Boussau B, Dutheil JY, Ranwez V (2012) Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. PLoS ONE **7:** e33852

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, et al (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res **18:** 65–76

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res **15:** 227–239

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature **463:** 178–183

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA **108:** 4069–4074

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115

Shang H, Li W, Zou C, Yuan Y (2013) Analyses of the NAC transcription factor gene family in *Gossypium raimondii* Ulbr.: chromosomal location, structure, phylogeny, and expression patterns. J Integr Plant Biol **55:** 663–676

Shiu SH, Bleecker AB (2001) Plant receptor-like kinase gene family: diversity, function, and signaling. Sci STKE **2001:** re22

Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KFX, Li WH (2004) Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. Plant Cell **16:** 1220–1234

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot **96:** 336–348

Soltis DE, Burleigh JG (2009) Surviving the K-T mass extinction: new perspectives of polyploidization in angiosperms. Proc Natl Acad Sci USA **106:** 5455–5456

Song X, Guo P, Li C, Liu CM (2010) The cysteine pairs in CLV2 are not necessary for sensing the CLV3 peptide in shoot and root meristems. J Integr Plant Biol **52:** 774–781

Sun W, Cao Y, Jansen Labby K, Bittel P, Boller T, Bent AF (2012) Probing the *Arabidopsis* flagellin receptor: FLS2-FLS2 association and the contributions of specific domains to signaling function. Plant Cell **24:** 1096–1113

Tan S, Wang D, Ding J, Tian D, Zhang X, Yang S (2011) Adaptive evolution of *Xa21* homologs in Gramineae. Genetica **139:** 1465–1475

Tang H, Bowers JE, Wang X, Paterson AH (2010a) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Natl Acad Sci USA **107:** 472–477

Tang P, Zhang Y, Sun X, Tian D, Yang S, Ding J (2010b) Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in four plant species. Plant Sci **179:** 399–406

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485:** 635–641

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596–1604

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, et al (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol **30:** 83–89

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al (2010) The genome of the domesticated apple (Malus × domestica Borkh.). Nat Genet **42:** 833–839

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet **45:** 487–494

Vetter MM, Kronholm I, He F, Häweker H, Reymond M, Bergelson J, Robatzek S, de Meaux J (2012) Flagellin perception varies quantitatively in *Arabidopsis thaliana* and its relatives. Mol Biol Evol **29:** 1655–1667

Wang GL, Ruan DL, Song WY, Sideris S, Chen L, Pi LY, Zhang S, Zhang Z, Fauquet C, Gaut BS, et al (1998) *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell **10:** 765–779

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, et al (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet **44:** 1098–1103

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet **43:** 1035–1039

Wu HJ, Zhang Z, Wang JY, Oh DH, Dassanayake M, Liu B, Huang Q, Sun HX, Xia R, Wu Y, et al (2012) Insights into salt tolerance from the genome of *Thellungiella salsuginea*. Proc Natl Acad Sci USA **109:** 12219–12224

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, et al (2011) Genome sequence and analysis of the tuber crop potato. Nature **475:** 189–195

Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, ÓMáille P, Osbourn A, Qi X (2012) Divergent evolution of oxidosqualene cyclases in plants. New Phytol **193:** 1022–1038

Yang T, Chaudhuri S, Yang L, Du L, Poovaiah BW (2010) A calcium/calmodulin-regulated member of the receptor-like kinase family confers cold tolerance in plants. J Biol Chem **285:** 7119–7126

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24:** 1586–1591

Yang Z, Wang Y, Zhou Y, Gao Q, Zhang E, Zhu L, Hu Y, Xu C (2013a) Evolution of land plant genes encoding L-Ala-D/L-Glu epimerases (AEEs) via horizontal gene transfer and positive selection. BMC Plant Biol **13:** 34

Yang ZL, Liu HJ, Wang XR, Zeng QY (2013b) Molecular evolution and expression divergence of the *Populus* polygalacturonase supergene family shed light on the evolution of increasingly complex organs in plants. New Phytol **197:** 1353–1365

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, et al (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature **480:** 520–524

Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79–92

Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, et al (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotechnol **30:** 549–554

Zhang XS, Choi JH, Heinz J, Chetty CS (2006) Domain-specific positive selection contributes to the evolution of *Arabidopsis* leucine-rich repeat receptor-like kinase (LRR RLK) genes. J Mol Evol **63:** 612–621

Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu SH (2009) Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. Plant Physiol **151:** 3–15